Plant Physiology®

Update

Using precision phenotyping to inform de novo domestication

Alisdair R. Fernie [iD] [1,2,*,†] **Saleh Alseekh** [iD] [1,2] **Jie Liu** [iD] [3] **and Jianbing Yan** [3,†]

1   Max Planck Institute for Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany
2   Centre of Plant Systems Biology and Biotechnology, 4000 Plovdiv, Bulgaria
3   National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, 430070 Wuhan, Hubei, China

*Author for communication: fernie@mpimp-golm.mpg.de (A.R.F.)
†Senior authors.
A.R.F. and J.Y. designed the review concept. All authors wrote the manuscript and finally approved the final version of the manuscript.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/general-instructions) are: Alisdair R. Fernie (fernie@mpimp-golm.mpg.de) and Jianbing Yan (yjianbing@mail.hzau.edu.cn).

Introduction

It is an oft-cited fact that toward the middle of this century, the rate of increase in global population is likely to outstrip that in global agricultural production. Moreover, productivity varies greatly across the globe, yet the majority of the burden on agriculture is placed on the cultivation of a small number of species largely in locations different from their origin of domestication and often subject to far different environmental conditions (Fernie and Yan, 2019). Recent technical developments—mainly the enhanced accessibility and affordability of next-generation sequencing technologies—have allowed the identification of over 100 domestication genes (Fernie and Yan, 2019). Many of these, such as those associated with the loss of shattering, seed size, and dormancy, are conserved across our crop species (Gross and Olsen, 2010; Lenser and Theissen, 2013). However, others seem to be specific to certain crops or crop types such as the modification of fruit shape (Xiao et al., 2008) or the evolution of tubers (Cheng et al., 2016; Hardigan et al., 2017). Having identified the genes, they can be used for de novo domestication, that is the genetic improvement of little cultivated or as yet undomesticated species. Key to this is the identification of species that display specific desired properties, for example higher production and fertilizer use

ADVANCES BOX

- Recent years have seen an explosion in the number of domestication alleles that have been demonstrated to underlie important agronomic traits.
- Gene editing techniques allow the manipulation of (multiple) such alleles into as yet undomesticated species as a method to produce crops better suited to a given environmental niche.
- Precision phenotyping approaches are beginning to be used in the design and execution of such projects and will undoubtedly become a key aspect of such endeavors.

efficiency, more balanced nutritional properties, better flavor or, alternatively, that are better suited for growth under certain agronomic conditions. To date, a handful of examples of de novo domestication have been published, which act as a proof-of-concept of the approach (Lemmon et al.,

Open Access

2018; Li et al., 2018; Zsögön et al., 2018; Hoyos et al., 2020; Yu et al., 2021). Moreover, several projects such as The African Orphan Crops Consortium (Jamnadass et al., 2020) have been initiated, which aim to use this approach at a very large scale. This Update focuses on recent progress made in both de novo domestication per se and in the use of precision phenotyping to assess suitable candidates for the process. Largely, but not exclusively, using teosinte-maize (Zea mays) as a case-study we (i) detail how computational methods can be used for identifying domestication genes and evaluating changes associated with these genes at both the phenotypic and molecular levels (Broman et al., 2019), (ii) demonstrate the power of precision phenotyping in evaluating the potential phenotypic scope of such interventions (Yang et al., 2020), (iii) describe the combined power of bioinformatics approaches and CRISPR/Cas libraries for streamlining agricultural engineering strategies (Childs et al., 2012; Liu et al., 2020), (iv) explain the benefit of harnessing recently acquired information concerning natural structural variations in crop and close-relative genomes (Alonge et al., 2020; Domínguez et al., 2020), (v) outline the potential role single cell sequencing will play in applying these methods to polyploid species (Luo et al., 2020), and finally (vi) discuss the prospects of recent approaches integrating molecular markers into metabolic models to improve genome selection (Tong et al., 2020). We discuss in detail the concepts underlying these approaches rather than the technical detail of either next-generation sequencing, genome editing, or plant phenotyping studies, which have been the subjects of recent reviews in their own right (Schneeberger and Weigel, 2011; Pacher and Puchta, 2017; Yang et al., 2020).

De novo domestication—the state of the art

The combination of global climate change, our expanding population and the increasing growth of plants for biofuel and other non-food purposes is leading to an ever-increasing demand for agronomic efficiency. Indeed, productivity varies greatly across the globe and while the production efficiency of cereals has kept pace with the human population, this is not true for all crops and some are even poorly suited to their area of cultivation. For example, the yields of cassava (Manihot esculenta) are three times higher in South East Asia than in Africa (Sonnewald et al., 2020). This instability aside, we only cultivate around 150 species to a large extent and 70% of the calories consumed by humans comes from only 15 of these species (Fernie and Yan, 2019). That said, more than 7,000 of the 400,000 extant plant species are regarded as semi-cultivated (Smýkal et al., 2018) and could represent important germplasm for the design of future crops. Here, we describe already published examples of plant de novo domestication before briefly outlining what makes a good candidate species for this approach. Following this introduction, we detail the integral role we envisage computational and precision phenotyping will play in future strategies of de novo domestication and re-domestication.

Four recent studies in the Solanaceae demonstrate the potential of gene editing for de novo domestication. Two of the studies included the orphan crop groundcherry (Physalis pruinosa; Lemmon et al., 2018; Kwon et al., 2020), while three included de novo domestication of the wild tomato Solanum pimpinellifolium or alteration of the MicroTom cutivar (Li et al., 2018; Zsögön et al., 2018; Kwon et al., 2020). The studies on groundcherry included development of a transformation procedure alongside the development of genomic resources for the species. Following the acquisition of this competence, the authors first knocked out SELF-PRUNING (SP) and SELF-PRUNING-5G (SP-5G) with the latter resulting in enhanced auxiliary flowering and increased fruit density, as well as targeting the CLAVATA (CV) pathway, which increased floral meristem size and led to additional floral organs (Lemmon et al., 2018). Analogous studies in S. pimpinellifolium targeted these three, and a further eight, genes (Li et al., 2018; Zsögön et al., 2018). They also proved successful in altering target traits, for example, doubling the yield of S. pimpinellifolium and enhancing lycopene levels five-fold. In the fourth study, both the MicroTom cultivar of S. lycopersicum and groundcherry were modified either by stacking a gene for tomato stem length with SP5G or SP or targeting the same stem length regulator alone. Both approaches led to a compact stature and early yielding plants suitable for urban agriculture (Kwon et al., 2020). Although we note that the work on MicroTom is clearly not an example of de novo domestication per se that on groundcherry is. Actually, the term de novo domestication is not universally accepted, given that domestication over millennia is acting on the whole genome in concert (Stetter et al., 2018; Stetter, 2020). While understanding these concerns, it is a term that has been rapidly adopted and a clear consensus concerning its meaning has been reached. In essence, de novo domestication refers to a deliberate modification of the sequence of a domestication gene in a lesser grown species with the aim of producing an agriculturally useful plant.

Although many of the early examples of de novo domestication have been carried out in the Solanaceae, the approach is by no means limited to this family, with examples in pennycress (Thlaspi arvense L.; McGinn et al., 2019), sunflower (Helianthus annuus; Ekar et al., 2019; Van Tassel et al., 2020), and the legume Vigna stipulacea (Takahashi et al., 2019). Pennycress has been de novo domesticated as a seed cover crop for the winter fallow period. For this purpose, CRISPR was used to produce insertion/deletion (indel) mutations in the FATTY ACID ELONGATION1 (FAE1) gene, thereby abolishing erucic acid production and creating an edible seed oil comparable to that of canola (McGinn et al., 2019). By contrast, conventional breeding has been used to domesticate the oilseed Silphium, resulting in increased aboveground biomass at the seedling and adult stages and a greater increase in seed yield, combining to a modest improvement in harvest index (Van Tassel et al., 2020). A similar approach has been utilized to follow sunflower domestication in order to develop a perenial crop that can

produce both high value vegetable oil and continuous ground cover (Ekar et al., 2019). Using an alternate approach, Takahashi et al. (2019) carried out ethyl methanesulfonate (EMS) mutagenesis and screening of a *V. stipulacea* population, isolating mutants with reduced seed dormancy and shattering, respectively. *Vigna* represents an interesting legume species for de novo domestication given that it exhibits rapid growth, a short vegetative stage, and broad resistance to pests and diseases. Although it will be quite time-consuming, the authors postulate that by pyramiding these mutant phenotypes, they will be able to generate a primitive crop which can be cultivated without pesticide. In addition to these specific examples, considerable research effort is also ongoing in close relatives of sorghum (*Sorghum bicolor*) and sugarcane (*Saccharum officinarum*; Paterson et al., 1995a, 1995b; Zhang et al., 2018) and the cases for utilizing many other crop wild relatives not least for their enhanced stress tolerance have been convincingly made (Zhang et al., 2018; Fernie and Yan, 2019).

A recent tour-de-force of the de novo domestication of the allotetraploid wild rice *Oryza alta* was published in Cell (Yu et al., 2021). In this article, a breeding route was presented in order to harness the advantages polyploidization in terms of genome buffering, vigorousness, and environmental robustness to rice. To do so, the authors developed an efficient transformation system, thus facilitating gene editing and a high-quality genome assembly of *O. alta*. In this example, *O. alta* was chosen after evaluating three wild species (8 *O. alta*, 2 *Oryza grandiglumis*, and 18 *Oryza latifolia* lines) and selecting *O. alta* due to its callus induction and regeneration properties. Following this, as a case study, six agronomically important traits, namely shattering, awn length, hull color, pericarp color, panicle shape, and grain width, were rapidly improved in *O. alta*, thereby demonstrating the feasibility of its de novo domestication. This study is arguably the first true de novo domestication and illustrates the importance of developing transformation systems and genome sequences as enabling steps for this process.

Accelerating the identification of domestication genes
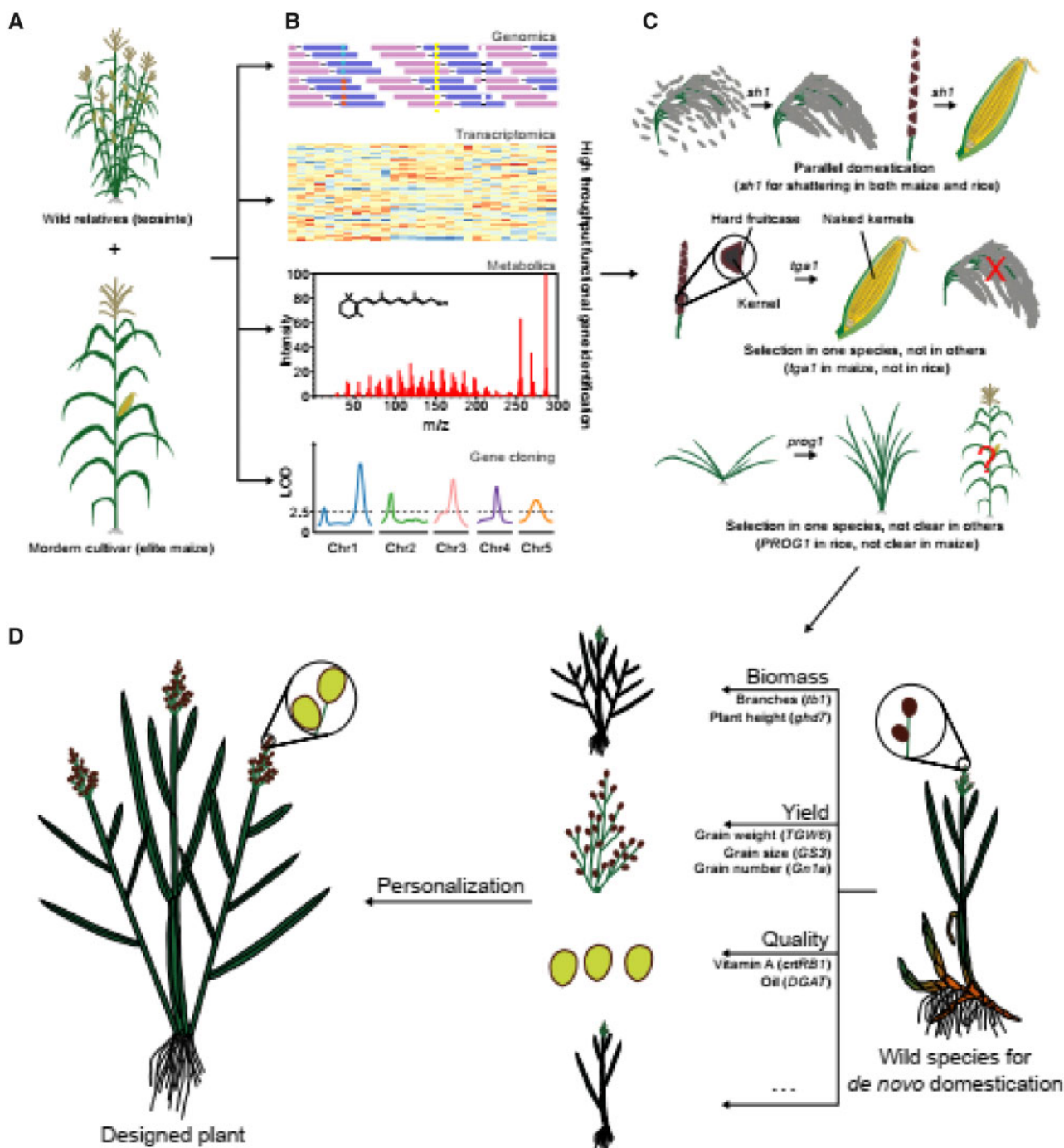
As stated above, rapid de novo domestication requires both genomic information and effective transformation procedures. Similarly, the alternative genetics-based approaches of crossing near relatives in attempt to introgress a trait of interest or the adoption of mutagenesis-based strategies have been greatly facilitated by the generation of computational tools that render them considerably easier (Huang and George, 2011; Youens-Clark et al., 2011; Meng et al., 2015; Broman et al., 2019; Wei et al., 2021). However, given that all of these have become relatively standard laboratory techniques, we discuss only the utility of genome data in the identification of domestication genes here.

The use of multiomics in de novo domestication is summarized in Figure 1, which we first describe here before dissecting the various layers involved in detail in the following paragraphs. Wild relatives and modern cultivars as exemplified by teosinte and maize (a comparison we routinely use, since, as mentioned above, it arguably represents the best characterized material covering plant domestication and improvement; Figure 1, A) are highly important for tracking signals of selection and thereby defining the genomic regions underlying the phenotypic changes during the process. Comparisons of the progeny of a biparental cross between the wild and cultivated species are commonly made, subjected to metabolomics and transcriptomics, and the underlying quantitative trait loci including domestication genes are cloned (Figure 1, B). Literature information concerning the roles of these genes in other species can then be mined, since where they were able to provide clear answers; previous studies have revealed that such genes are either subject to parallel domestication or species-specific domestication events (Figure 1, C). Following such assessments traits of interest, such as biomass, yield, and quality, can be modified via gene editing of the underlying genes allowing de novo domestication and personalization of the de novo domesticate (Figure 1, D).

In practice, the tracking of signals of selection is carried out by computational analyses known as selective sweeps, which are essentially a measure of the genetic diversity between cultivated and ancestral genotypes. Following this approach, it is assumed that those regions that have been selected for during domestication harboring less diversity than those that regions that are subject to less stringent selection (Zhu et al., 2018). When a genetic variant increases its frequency due to positive selection, the adjacent alleles also increase in their frequency in a process termed hitchhiking (Smith and Haigh, 2007). When the genetic variant under selection reaches high frequency (or even fixation), the hitchhiking effect reduces (or even eliminates), the genetic variation around the selected locus in an effect known as a selective sweep (Pavlidis and Alachiotis, 2017). Local F_{ST} outlier tests can be used to detect selective sweeps between populations of wild and domesticated taxa (Gepts, 2014). There are many different methods based on the F_{ST} statistic, which largely vary in the underlying model used to calculate the null distribution of F_{ST} values (see e.g. Bonhomme et al., 2010; Lotterhos and Whitlock, 2015). Selecting a suitable method is complicated, with considerations including the sampling scheme taken, the total size of the dataset, intensity of selective pressure, and the genetic structure of the population all needing to be taken into account (Barrera-Redondo et al., 2020). Illustrative examples of the use of these tests in plants include the detection of domestication genes involved in apple (*Malus domestica*) fruit development, size, acidity and sugar content (Khan et al., 2014), size, color, and disease resistance in tomato (Zhu et al., 2018), and those for oil biosynthesis in sunflower (Baute et al., 2015).

Given the torrent of genome sequence information, it seems likely that optimization of computational approaches will greatly accelerate identification of domestication genes.

Figure 1 Overview of knowledge-driven de novo domestication. A, The wild relatives and modern cultivars are vital germplasms (e.g. teosinte and maize) for tracing selection signals during plant domestication and also functional genomics research. B, Multiple-omics, including genomics, transcriptomics, metabolomics, and phenomics, are foundations for high-throughput gene cloning and gene function analysis. C and D, Wild species chosen for de novo domestication could be quickly shaped to meet different kinds of demands through selections on detected signals during previous domestications, especially parallel selected genes. Ideal plants can thus be designed through the pyramiding and modification of genes of diverse function.

Indeed, if pedigree is established from genome information and this allows clear discrimination of crop wild relatives, landraces, and cultivars, similar experiments can be carried out to define domestication and improvement genes for many species. It is worth noting that a mere 15 years ago the number of confirmed domestication genes numbered just over two dozen (Doebley et al., 2006); however, there are now well over a hundred with notable recent genes being identified for dormancy in rice, soybean (*Glycine max*), and tomato (Wang et al., 2018), the evolution of tubers in for example potato (Hardigan et al., 2017) and grain filling in maize (Sosso et al., 2015). Many of the genes involved are

transcription factors (Fernie and Yan, 2019) as would perhaps be anticipated given the dramatic upregulation of gene expression on domestication evidenced in teosinte to maize comparisons (Swanson-Wagner et al., 2012). Beyond searching genomes for signatures of selection, detailed characterization of the transcriptome, metabolome, and phenome will represent important information when choosing which species is most appropriate for de novo domestication. We discuss the first two of these here, but the third is so important that it warrants its own section below. Before we do so, however, it should be noted that the aim of all of these is a broad characterization of lesser studied species rather than merely an assessment of the variance of a trait of interest. Put another way, the domestication genes already described are the likely tools for de novo domestication but deciding the likely species to domesticate first requires a better understanding of the physiology of a range of species. One could, for example, look for species that thrive in a particular environment, have a nutritionally excellent chemical composition, high nitrogen or water use efficiencies, or are naturally pest-resilient. That said, we certainly appreciate that de novo domestication will be easier for similar species, such as the excellent examples in rice and the Solanaceae. A major question for the future is how easy it will be to transfer allelic diversity between members of different genera, tribes, or even families where there are as yet no domesticated species. It is too early to answer this question. Many domestication traits are shared by a very wide number of species, while others are more taxonomically restricted (Meyer et al., 2012). This clearly renders de novo domestication of completely wild species that are taxonomically distant from our major crops exceptionally hard. By contrast, those traits which are widely shared represent obvious targets; however, that said, even if de novo domestication is restricted to crop wild relatives, it is likely to allow great agronomic advances.

Transcript profiles are tissue-specific and subject to temporal control, thus transcriptomics can reveal important loci involved in domestication traits such as suppression of secondary metabolites, changes in form, size, taste, absence of defense mechanisms, seed dormancy, and many other traits (Barrera-Redondo et al., 2020). For example, the assessment of altered gene expression on domestication was achieved in maize, tomato, lettuce (*Lactuca sativa*), and common bean (*Phaseolus vulgaris*) by comparing the total gene expression of a range of wild species, progenitors, and cultivars, respectively (Swanson-Wagner et al., 2012; Koenig et al., 2013; Bellucci et al., 2014). A comparison of 38 diverse maize and 24 teosinte provided evidence for more than 600 genes having significantly different expression levels and almost twice as many altered co-expression patterns (Swanson-Wagner et al., 2012). Included among the 600 were a mere 46 that had been previously identified as targets of selection and a similar number presumed to result as an effect of inbreeding.

Similarly to the above-described maize study, research in tomato used RNA-seq to define both gene sequence and expression divergence between cultivated tomato and five related wild species (Koenig et al., 2013). Based on sequence differences, Koenig et al. (2013) detected footprints of positive selection in over 50 genes and documented thousands of shifts at the gene-expression level. These rapidly evolving genes are commonly associated with environmental responses and stress tolerance. Similarly, decreased nucleotide and expression diversity and modified co-expression patterns characterized the domestication of common bean (Bellucci et al., 2014). Moreover, RNA sequencing of 240 accessions of lettuce revealed a list of regions as putative selective sweeps that occurred during domestication and divergence, respectively (Zhang et al., 2017). Genome-wide association studies (GWASs) further identified 5,311 expression quantitative trait loci (eQTL) regulating the expression of 4,105 genes, including nine eQTLs regulating genes associated with flavonoid biosynthesis and thereby underlying color and nutritional content of the crop. Indeed, the GWAS approach is highly useful alongside the selective sweep approach in characterizing changes on domestication. While there are variants as to the mathematical model underlying such analyses, they are all essentially highly similar in that they associate genotypic and phenotypic variance (lists of GWAS databases and commonly used approaches are provided in Tables 1, 2).

A wide range of bitter-tasting compounds were selected against during the processes of domestication and improvement including the steroidal glycoalkaloids of tomato and potato (*Solanum tuberosum*; Itkin et al., 2013; Schwahn et al., 2014), β-L-oxayl-2,3-diaminopropionic acid (β-L-ODAP) in grass pea (Emmrich et al., 2019), curcurbitadienol in cucumber (*Cucumis sativus*; Shang et al., 2014; Zhou et al., 2016), glucosinolates in broccoli (*Brassica oleracea*; Drewnowski and Gomez-Carneros, 2000), and the flavone-7-*O*-neohesperidoside in citrus (Frydman et al., 2013). While the decrease in the levels of these metabolites is rather predictable and moreover easy to rationalize metabolome wide studies in wheat (Beleggia et al., 2016), maize (Xu et al., 2019), rice (Deng et al., 2020), and tomato (Zhu et al., 2018) and latterly in lettuce (Zhang et al., 2020) and tea (Zhang et al., 2020) revealed far more complex changes.

In the first of these studies, Beleggia et al. (2016) showed that the primary domestication of wheat was characterized by a reduction in unsaturated fatty acids on the primary domestication with altered amino acid content characterizing the secondary domestication. Maize, by contrast was characterized by alkaloid, terpenoid, and lipid changes at the divergence between teosinte and tropical maize, whereas benzoxazinoid levels changed at the divergence between tropical and temperate maize (Xu et al., 2019) and rice displayed different changes again (Deng et al., 2020). Tomato domestication had highly diverse effects on the metabolome, with many metabolic changes being associated with the increase in size, others with breeding for color preferences and yet further by the introgression of disease resistance from wild relatives (Zhu et al., 2018). Likewise in lettuce, quinate and chlorogenic acid levels were strongly reduced on domestication of lettuce, probably as a

Table 1 List of database and tools commonly used for GWAS and genomic studies

Name	Link	Refrence
easyGWAS: a cloud-based platform for comparing the results of genome-wide association studies	https://easygwas.ethz.ch/	Grimm et al. (2017)
Matapax: an online high-throughput genome-wide association study pipeline	https://matapax.mpimp-golm.mpg.de/	Childs et al. (2012)
GWAPP: a web application for genome-wide association mapping in Arabidopsis	https://gwas.gmi.oeaw.ac.at/	Seren et al. (2012)
GWAS Atlas: a curated resource of genome-wide variant–trait associations in plants and animals	https://bigd.big.ac.cn/gwas/	Ye et al. (2020)
PGSB PlantsDB: updates to the database framework for comparative plant genome research	https://pgsb.helmholtz-muenchen.de/plant/index.jsp	Spannagl et al. (2016)
Phenotypic and genome-wide association with the local environment of Arabidopsis	http://www.personal.psu.edu/sma3/CLIMtools.html	Ferrero-Serrano and Assmann (2019)
RiceVarMap: a comprehensive database of rice genomic variations	http://ricevarmap.ncpgr.cn/v2/	Zhao et al. (2014)
GWASpro: a high-performance genome-wide association analysis server	https://bioinfo.noble.org/GWASPRO/	Kim et al. (2018)
MaizeGDB 2018: the maize multi-genome genetics and genomics database	https://www.maizegdb.org/	Kim et al. (2018)
TASUKE+: a web-based platform for exploring GWAS results and large-scale resequencing data	https://tasuke.dna.affrc.go.jp/	Kumagai et al. (2019)
ZEAMAP, a comprehensive database adapted to the maize multi-omics era	http://www.zeamap.com	Gui et al. (2020)
MaizeCUBIC: a comprehensive variation database for a maize synthetic population	http://modem.hzau.edu.cn/	Luo et al. (2020)
CARMO: a comprehensive annotation platform for functional exploration of rice multi-omics data	http://bioinfo.sibs.ac.cn/carmo	Wang et al. (2015)
AraQTL—workbench and archive for systems genetics in Arabidopsis thaliana	http://www.bioinformatics.nl/AraQTL/	Nijveen et al. (2017)
WheatExp: an RNA-seq expression database for polyploid wheat	http://wheat.pw.usda.gov/WheatExp/	Pearce et al. (2015)
CerealsDB 2.0: an integrated resource for plant breeders and scientists	https://www.cerealsdb.uk.net/cerealgenomics/CerealsDB/indexNEW.php	Wilkinson et al. (2012)
The Triticeae Toolbox: combining phenotype and genotype data to advance small-grains breeding	http://triticaetoolbox.org	Blake et al. (2016)
The AraGWAS Catalog: a curated and standardized Arabidopsis thaliana GWAS catalog	https://aragwas.1001genomes.org	TogninalLi et al. (2018)
Gramene: a resource for comparative analysis of plants genomes and pathways	http://www.gramene.org	Tello-Ruiz et al. (2018)
Ensembl Genomes 2020—enabling non-vertebrate genomic research	http://www.ensemblgenomes.org	Howe et al. (2020)
SnpHub: an easy-to-set-up web server framework for exploring large-scale genomic variation data in the post-genomic era with applications in wheat	http://wheat.cau.edu.cn/Wheat_SnpHub_Portal/	Guo et al. (2020)

consequence of the desire to reduce bitterness (Zhang et al., 2020). By contrast tea, probably as a result of its complex domestication, did not display clear changes in its metabolite content across the domestication and improvement processes (Zhang et al., 2020).

While transcriptomics and metabolomics and corresponding GWAS analysis can clearly allow the computational discrimination of domestication genes, the number and scale of changes at the transcript and metabolite level suggest that transcriptomics and metabolomics will be necessary for the regulatory control of any future de novo domesticated product. Such control has been debated for CRISPR/Cas lines (Fedorova and Herman, 2020; Fraser et al., 2020) in general but will be particularly important for de novo domestication events (either using CRISPR/Cas or conventional breeding) using species that are currently not eaten in great amounts in order to ensure their

safety. For this purpose, a comparison of the metabolomes of de novo domestication events against conventional crops will prove highly worthwhile, although it will be important to improve the curation of such a comparison on our current knowledge concerning the (anti)nutritional values of the individual metabolites in our major crops such that this is as good as the FAO yield statistics (http://www.fao.org).

The use of precision phenotyping to demonstrate the phenotypic scope of de novo domestication

As demonstrated above, a pipeline for identifying target genes for de novo domestication is relatively easy following analysis of our current crops and their wild relatives. Indeed, a number of key genes for improvement in orphan crops

Table 2 Commonly used packages for conducting GWAS[a]

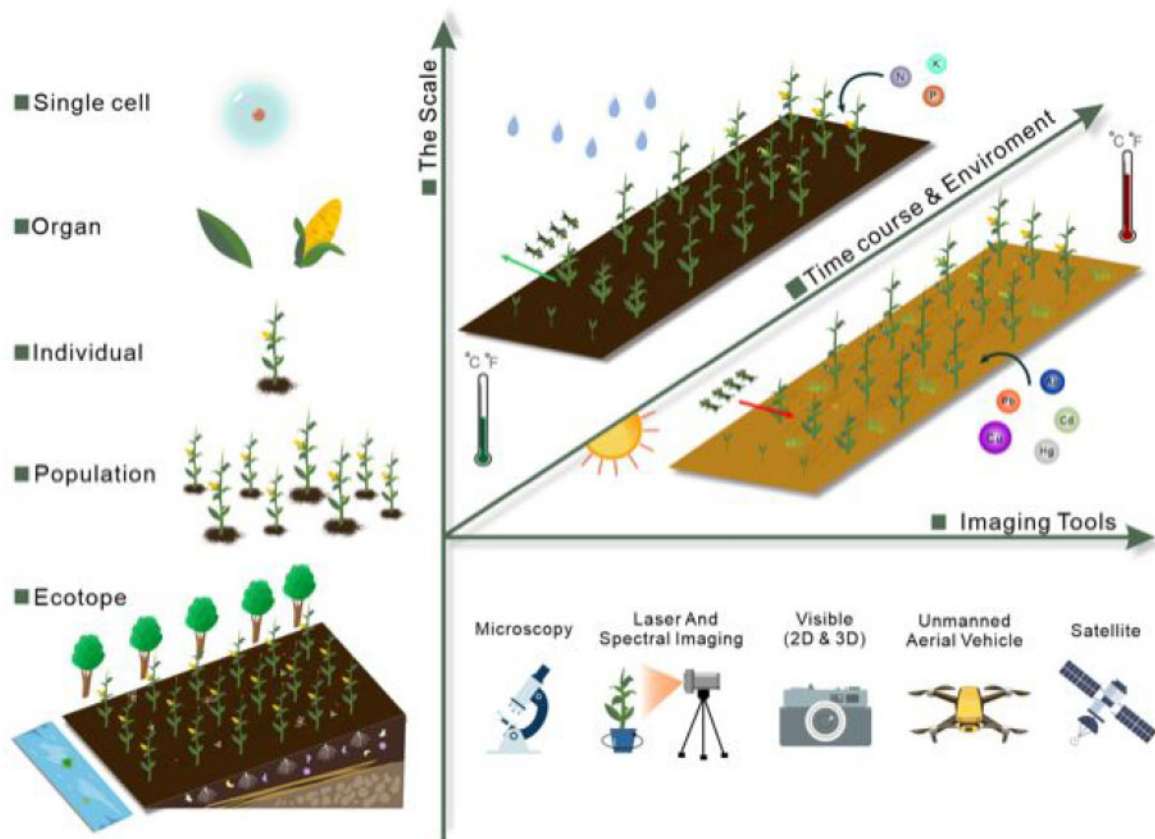| Package	Description	Web page
TASSEL	Variety of algorithms MLM, GLM, weighted MLM, genomic selection, fast association; supports P3D compression; can process GBS data; designed to determine dominance/additivity of effects; user-friendly GUI	http://www.maizegenetics.net/tassel
GAPIT	Package that can perform MLM and EMMA; supports P3D and EMMAx; works via R language	http://www.maizegenetics.net/gapit
A-DTest	R package is ADGWAS: for GWAS	https://github.com/maizego/A-D-test
EMMAX	Efficient mixed-model method for large genomic datasets; command-line interface only	http://genetics.cs.ucla.edu/emmax/index.html
GEMMA	Standard/multivariate/Bayesian linear mixed-model framework; estimates quantitative genetic traits and proportioning of variance; command-line interface only	http://www.xzlab.org/software.html
ANGSD	Useful when genotypic states are not known with certainty; measures population genetic parameters; command-line interface only	http://www.popgen.dk/angsd/index.php/ANGSD
Plink	Wide-ranging toolset for conducting GWAS; originally designed for human genome data; command-line interface only	http://pngu.mgh.harvard.edu/purcell/plink/
Lrgpr	Allows for testing of G × G and G × E; works via R language	http://lrgpr.r-forge.r-project.org/

aModified from Burghardt et al. (2017).

have previously been discussed, including genes involved in plant architecture (barley *SEMI-DWARF1*, *TEOSINTE BRANCHED1*, *DEEPER ROOTING1*, *PHOSPHORUS-STARVATION TOLERANCE1*, *PROG1*; Jia et al., 2009; Uga et al., 2013; Mai et al., 2014; Shang et al., 2014; Studer et al., 2017), seasonal flowering time (*PHOTOPERIOD-H1*, *CENTRORADIALIS*; Turner et al., 2005; Comadran et al., 2012), light competition (*PHYB1* and *PHYB2*; Sheehan et al., 2007), seed or fruit retention (*SHATTERING1* and others; Lin et al., 2012; Meyer and Purugganan, 2013), fruit size (*FRUIT WEIGHT2.2*; Frary et al., 2000), and length of the juvenile stage (*TERMINAL FLOWER 1*, *FLOWERING LOCUS T*; Bergonzi and Albani, 2011; Yamagishi et al., 2014). Similarly, as the proof-of-concept studies described above demonstrate, targeted manipulation of multiple genes can be readily carried out. That said, selecting which relative is the best choice to target for the process is considerably more difficult.

We believe that this question is best resolved by comparative phenotyping of a range of species of local wild species against the major cultivated crop as well as select genotyping of the level of allelic variance in the domestication gene in question. To the best of our knowledge, such experiments, while perhaps underway, have not yet been published. We did, in a previous article, suggest a range of noncultivated and semi-cultivated relatives (Fernie and Yan, 2019) and this list still stands. However, it is important to note that not only are these species under-utilized but they are also understudied. Indeed, the same could be held true until recently even for major crops such as cassava, sweet potato, and yam as well as nutritionally important crops such as quinoa. The recent publication of the genomes of these species (Wang et al., 2014; Jarvis et al., 2017; Yang et al., 2017; Scarcelli et al., 2019) alongside considerably better characterization of them at the physiological and metabolic levels (Obata et al., 2020; Price et al., 2020; Sonnewald et al., 2020) provides an effective blueprint as to how candidates for de novo domestication should be assessed. We summarize the set of tools that we believe should be brought to bear in field trials comparing the effects of established crops and candidates for de novo domestication in Figure 2. Here, we suggest that it will be important to use contemporary technology to consider all levels of the spatial hierarchy from single cells (Luo et al., 2020) to ecosystem models (Tian et al., 2020), using a broad range of imaging tools to record and also infer trait variances (Yang et al., 2020) In this vein, the power of inferential data has been provided by a recent proof of concept study comparing hyperspectral imaging of metabolite content to measurements made by mass-spectrometry (Vergara-Diaz et al., 2020) implying that this may shortly be as reliable as spectral measurements of photosynthesis from unmanned aerial vehicles have proven to be (Gago et al., 2020; Yang et al., 2020). Indeed, the data emanating from such imaging platforms (and those presented in Figure 3) will certainly form an essential component of deciding which species would be ideal for any given environment for the purposes of de novo domestication. As an extension to this strategy, the use of the reciprocal transplant strategy (Hereford, 2009; Ågren and Schemske, 2012; Sork, 2018), whereby species are planted and monitored in the environment of one another in addition to their own, would allow the adaptation of exotic species to be quantified in a human-controlled environment as a first indicator that aids in the selection of the best candidates for de novo domestication.

As we state above, to our knowledge, public data are not currently available for comparison of domesticated and nondomesticated species. However, all of these phenotyping methods have been demonstrated to provide reliable information (Watt et al., 2020; Yang et al., 2020). As such we would envisage the use of such precision phenotyping approaches, in comparative analyses, to be instrumental in narrowing down the lists mentioned above and beginning to define ideal targets for de novo domestication. This is very easy to state but how do we propose that this will work in practice. Well for starters examination of certain key parameters could be used for example the rate of

Figure 2 Utilizing multiple crop phenotyping tools for accurate high-throughput acquisition and analysis of multidimensional phenotypes on organism-wide scale (from single cell to ecotope). Examples range from micro to macro scale, through different environments (abiotic stress, biotic stress, etc.) and across the entire crop developmental process.

photosynthesis, ease of harvest, productivity, abiotic resilience, biotic resistance, ability to prosper in a dense stand, etc. Such traits could be computationally afforded different weighting with regard to the most challenging aspects of the environment the novel crop was planned for thereby providing a ranking of candidate species for a given environment. At a practical level, the complexity of obtaining the genome sequence and/or of generating allelic variants in the gene(s) of interest (via either gene editing or introgression) would likely considerably narrow these lists. However, despite current limitations, genome sequencing solutions for complex genomes are becoming available and as we detail below considerable advances are also being made in transformation of species that were previously regarded as recalcitrant.
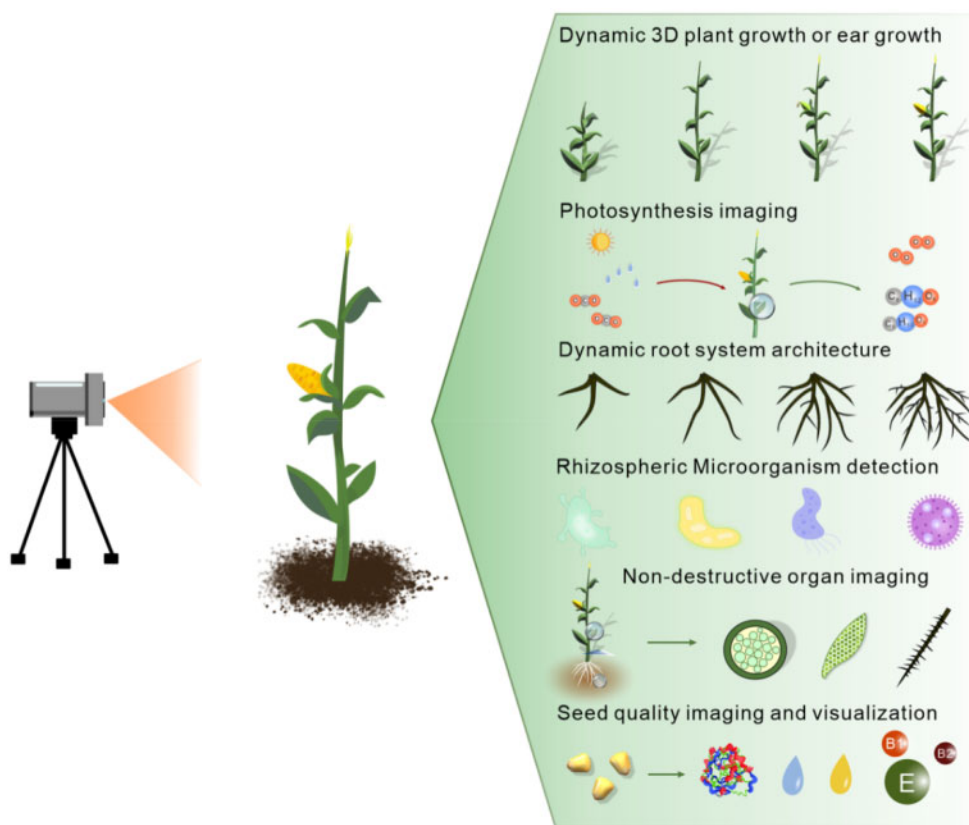
The power of CRISPR/Cas libraries

Following the selection of the candidate species for de novo domestication, two further hurdles remain. First, it would be prudent to have the genome of the species to be domesticated in hand such that one could be sure that the genetic interventions planned were sensible. Next-generation sequencing technologies have rendered this relatively facile and even large polyploidy genomes can be tackled without

trepidation. Second, on having identified the genetic targets to modify a route for genetic manipulation is needed. For this purpose, two clear possibilities exist. Using classical breeding, genes of interest can be introgressed into the candidate for de novo domestication providing that the source of these genes is closely genetically related to the recipient. Alternatively, chemical mutagenesis or the genome editing approach can be used. The use of the latter approach is complicated by the fact that regulatory concerns will render agricultural production of plants created by such technologies difficult in certain countries. However, given that it is more rapid, and more precise we will concentrate on this approach here. While Zinc Finger Nucleases and Transcription Activator like effector nucleases are also utilized, Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein (CRISPR-Cas) has moved to the fore primarily due to its simplicity and efficiency. A great advantage of the CRISPR/Cas system is that it can be used to edit multiple genes simultaneously (Xie and Yang, 2013; Zhang et al., 2014; Ma et al., 2015; Qi et al., 2016), and can generate large mutant libraries in a scalable high-throughput manner (Lu et al., 2017; Meng et al., 2017; Liu et al., 2020).

CRISPR-Cas systems have already shown their superiority in precision breeding through editing coding regions and

Figure 3 The utilization of crop phenotyping techniques. They can be applied widely for example in dynamic measurements of 3D plant growth, ear growth, photosynthesis, root system architecture, rhizospheric microorganism detection, non-destructive organ imaging, and the visualization of seed quality distribution.

knocking out large numbers of genes of interest (reviewed in Chen et al., 2019; Liu et al., 2021). However, fine-tuning of the expression of target genes via editing cis-regulatory sequences or changing the status of epigenetic marks is also very promising for future breeding designs (Liu et al., 2021). It is particularly worth mentioning that the further application of big data analysis methods utilizing machine-learning approaches may enable us to accurately understand and predict the function of each gene, segment of sequence, or even each base. Combined with CRISPR technology, not only we can accurately modify target sequences, but also can create new sequences, thereby altering their function(s). This will render future de novo domestication both more efficient and more accurate (Liu et al., 2020). In this vein, Rodríguez-Leal et al. (2017) provided a seminal example of obtaining quantitative variation in the important agronomic traits fruit size, inflorescence branching, and plant architecture via gene editing. While this was carried out in a domesticated species as a proof-of-concept, it is clearly conceivable that introducing such common domestication traits to underutilized species may well improve their prospects as agricultural commodities.

As mentioned above, a major constraint in our ability to harness the potential of other species is the difficulty in producing the allelic variance. While in many cases this goal could be achieved by mutagenesis or gene introgression via classical breeding, such approaches are relatively slow. Therefore, expanding the number of species that can be transformed is of vast importance. Recent studies have reported considerable improvements in the efficiency of plant regeneration from tissue culture are achievable by overexpression of plant developmental regulators including *LEAFY COTYLEDON1* or *LEAFY COTYLEDON2* (Lotan et al., 1998; Stone et al., 2001), *WUSCHEL* (Zuo et al., 2002), or *BABY BOOM* (Boutilier et al., 2002). Moreover, a recent study has shown that a fusion protein combining wheat *GROWTH-REGULATING FACTOR4* and its cofactor *GRF-INTERACTING FACTOR1* substantially increases the efficiency and speed of regeneration in wheat, triticale, and rice and increases the number of transformable wheat genotypes (Debernardi et al., 2020). These examples, alongside that of the transformation of the wild rice *O. alta* would appear to offer great promise for this approach. However, the ability to introduce the planned modification into the genome of the potential novel crop—be it by classical, mutagenesis, or gene editing approaches—should not be underestimated as it is currently the major challenge facing the de novo domestication approach. A couple of recent developments, in addition to those mentioned above offer hope here though. The first is the development of efficient grafting methods that would allow wild species to act as root-stocks as a route to novel agricultural products. This approach has

often been discarded due to graft incompatibilities, but this challenge has been overcome in many crops (Notaguchi, 2020). The second is the recent finding by the group of Dan Voytas that gene-edited dicots can be generated via de novo meristem induction (Maher et al., 2020). In this method, developmental regulators and gene editing reagents can be delivered to somatic cells in order to generate inheritable changes in sequence via a route that bypasses tissue culture. The development of such approaches thus provides optimism that in the future the challenge of modifying lesser grown plants could prove less daunting.

Despite the immense interest in these approaches, there are some concerns about the application of the CRISPR-Cas system in crop breeding, the biggest of these being the possibility of deleterious effects caused by the integration of transgenic constructs or off-target mutations. Several studies have documented the off-target effects of the CRISPR-Cas system in plants (Xie and Yang, 2013; Zhang et al., 2014; Endo et al., 2015; Jacobs et al., 2015; Jin et al., 2019). Computational analysis of the likelihood of off-target effects is a pre-requisite for precise de novo domestication approaches. In parallel, other methods of assessing genome edited crops such as at the level of the metabolome have been proposed as methods of ensuring that unintended effects of the editing can be monitored and, if these are negligible, that such crops can be regarded as safe (Fraser et al., 2020). Such analyses are important for all new crops; however, they will be far more important for true do novo domesticates as opposed to minor crops that are already consumed such as members of the amaranth genus (Stetter et al., 2020), or fonio millet (*Digitaria exilis*; Abrouk et al., 2020). It can only be hoped that a combination of such approaches, alongside the publicity generated by the 2020 Nobel Prize for Chemistry will allay public skepticism of gene editing.

One aspect that is difficult to envisage being tractable by gene editing is the harnessing of trait variation resulting from natural structural genome variants (Alonge et al., 2020; Alseekh et al., 2020; Domínguez et al., 2020; Fraser et al., 2020). It is becoming apparent from resequencing and the assembly of pan-genomes that structural variants beyond the mere addition or deletion of genes play important roles in shaping crop phenotypes. The results of two recent studies in tomato are particularly pertinent here that of Domínguez et al. (2020) revealed the importance of transpostional insertion on transcription with potential consequences on virus and Phytophthora resistance as well as shelf life while GWASs revealed structural variants associated to a number of these (Domínguez et al., 2020). Similarly, the study of Alonge et al. (2020) identified a P450 gene duplication underlying a fruit weight quantitative trait loci (which was previously thought to be associated with a single nucleotide polymorphism (SNP)) and other structural variants, which were required for breeding of the jointless trait (Alonge et al., 2020). Adoption of introgression-based strategies would be one current approach to harness such variation. However, whether such approaches will be necessary

awaits further, more detailed analysis of structural genome variants in the majority of our crop species. That said, it is important to note here that the CRISPR/Cas9 has already been demonstrated to display the diversity required to address these questions in plants (Li and Xia, 2020). In particular, it has been shown to induce chromosomal translocations (Beying et al., 2020), large inversions (Schwartz et al., 2020), and to change recombination patterns (Schmidt et al., 2020).

The prospects of integrating molecular markers and metabolites into models to improve phenotype prediction

Several other computational tools warrant discussion within the context of de novo domestication, including prediction of changes in the levels of molecular and morphological aspects of phenotype. A relatively simple example that is pertinent in this context is the recent attempt to model gene expression of the phenylpropanoid pathway a wide range of wild species tomato on the basis of metabolome data from the same samples (Tohge et al., 2020) with this being the culmination of several years of experience in integrative analyses reviewed in a previous Update in Plant Physiology (Tohge et al., 2015). Something that has proven considerably more complex is the prediction of yield from metabolomics data. This has nevertheless been attempted (Meyer et al., 2012; Riedelsheimer et al., 2012; Rosado-Souza et al., 2015). While generally speaking it is difficult to find a single metabolite whose level is predictive of yield, a wide number of studies have revealed metabolic signatures for this in a range of species including Arabidopsis, maize, wheat, and tomato (Schauer et al., 2006; Meyer et al., 2012; Riedelsheimer et al., 2012; Obata et al., 2015; Vergara-Diaz et al., 2020). Application of such tools to de novo domesticated crops will be a highly useful strategy to search for additional manipulations that will allow further improvement of these novel crops.

In addition, genomic selection approaches that utilize molecular markers and machine learning to identify superior genotypes with improved traits such as growth have started to incorporate –omics level data. Given that this topic has recently been comprehensively reviewed, we do not discuss it in detail here (Tong and Nikoloski, 2020). Suffice to say that results from a recent study suggest that integrating molecular markers into metabolic models can dramatically improve the prediction accuracy of genomic selection strategies (Tong et al., 2020). We thus feel it likely that the integration of such extended models, as well as the integration of other machine learning strategies such as those described in Liu et al. (2020) will most likely prove highly informative in the development of second generation de novo domesticates.

Conclusion and future perspectives

As we have described above, a growing number of examples have demonstrated that it has become relatively facile to de

novo domesticate plants. Indeed, technological advances in nucleotide sequencing and gene editing have greatly expanded our capability to tackle such ambitious projects. We argue here that given these stunning advances, one of the most difficult aspects of this procedure will be the choice of a suitable starting species presenting a potential route by which to address this question. Multiple roles for computation can be envisaged in the process of de novo domestication from comparative phenotypic evaluation of potential starting species, through genome analysis and population genomics to the comparative evaluation of de novo domesticates and their progenitors. The need for computation in the design of CRISPR/Cas strategies, the evaluation of GWASs, multi-omics integration, genomic prediction, and deep learning is equally pervasive. Indeed, with the increasing sophistication of available genotyping and phenotyping data, it is becoming harder and harder to imagine any large-scale de novo domestication program not employing the tools of precision phenotyping. Despite its great promise, a number of important and even fundamental questions remain as yet unanswered (see Outstanding Questions Box). Moreover, considerable challenges that limited and constrained crop domestication over millennium (Stetter, 2020) will ultimately also need to be considered when approaching de novo domestication. Among these the genetic architecture underlying traits, the level of standing genetic variation for domestication traits, and the accumulation of genetic load (i.e. deleterious genetic variants) have been noted as important factors determining the extent of domestication of various species. Less domesticated crops are often well adapted to diverse environments and of high nutritional value but need improvement in key domestication traits to render them serious alternatives to our existing crops. We believe that the key steps in de novo domestication are three-fold: (i) the comprehensive evaluation of a wide range of lesser grown crops alongside the acquisition of (ii) high-quality genome sequences, (iii) knowledge concerning functional genes, and (iv) the competence to generate variance in key genes of interest. While the examples to date suggest that such approaches are possible, a lot of research and development will be required before they become routine.

OUTSTANDING QUESTIONS BOX

- Where is our greatest need for de novo domestication?
- What are the most suitable species to act as progenitors to the de novo domesticates?
- Can we improve minor crops with the tools of de novo domestication?
- Can we tailor the generation of de novo domesticates to perform better in projected future climates?
- Will it be possible to utilize deep-learning methods to improve future strategies?

References

Abrouk M, Ahmed HI, Cubry P, Šimoníková D, Cauet S, Pailles Y, Bettgenhaeuser J, Gapa L, Scarcelli N, Couderc M, et al. (2020) Fonio millet genome unlocks African orphan crop diversity for agriculture in a changing climate. Nat Commun **11**: 4488

Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism *Arabidopsis thaliana* in its native range. New Phytol **194**: 1112–1122.

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. Cell **182**: 145–161.e123

Alseekh S, Scossa F, Fernie AR (2020) Mobile transposable elements shape plant genome diversity. Trends Plant Sci **25**: 1062–1064

Barrera-Redondo J, Piñero D, Eguiarte LE (2020) Genomic, transcriptomic and epigenomic tools to study the domestication of plants and animals: a field guide for beginners. Front Genet **11**: 742

Baute GJ, Kane NC, Grassa CJ, Lai Z, Rieseberg LH (2015) Genome scans reveal candidate domestication and improvement genes in cultivated sunflower, as well as post-domestication introgression with wild relatives. New Phytol **206**: 830–838

Beleggia R, Rau D, Laido G, Platani C, Nigro F, Fragasso M, De Vita P, Scossa F, Fernie AR, Nikoloski Z, et al. (2016) Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. Mol Biol Evol **33**: 1740–1753

Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, Minio A, Rau D, Rodríguez M, Panziera A, et al. (2014) Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. Plant Cell **26**: 1901–1912

Bergonzi S, Albani MC (2011) Reproductive competence from an annual and a perennial perspective. J Exp Bot **62**: 4415–4422

Beying N, Schmidt C, Pacher M, Houben A, Puchta H (2020) CRISPR-Cas9-mediated induction of heritable chromosomal translocations in Arabidopsis. Nat Plants **6**: 638–645

Blake VC, Birkett C, Matthews DE, Hane DL, Bradbury P, Jannink JL (2016) The Triticeae toolbox: combining phenotype and genotype data to advance small-grains breeding. Plant Genome **9**

Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, Sancristobal M (2010) Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics **186**: 241–262

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AA, Miki BL, et al. (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. Plant Cell **14**: 1737–1749

Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA (2019) R/qtl2: software for mapping

quantitative trait loci with high-dimensional data and multiparent populations. Genetics **211**: 495–502

Burghardt LT, Young ND, Tiffin P (2017) A guide to genome-wide association mapping in plants. Curr Protoc Plant Biol **2**: 22–38

Chen K, Wang Y, Zhang R, Zhang H, Gao C (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. Annu Rev Plant Biol **70**: 667–697

Cheng F, Sun R, Hou X, Zheng H, Zhang F, Zhang Y, Liu B, Liang J, Zhuang M, Liu Y, et al. (2016) Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. Nat Genet **48**: 1218–1224

Childs LH, Lisec J, Walther D (2012) Matapax: an online high-throughput genome-wide association study pipeline. Plant Physiol **158**: 1534–1541

Comadran J, Kilian B, Russell J, Ramsay L, Stein N, Ganal M, Shaw P, Bayer M, Thomas W, Marshall D, et al. (2012) Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. Nat Genet **44**: 1388–1392

Debernardi JM, Tricoli DM, Ercoli MF, Hayta S, Ronald P, Palatnik JF, Dubcovsky J (2020) A GRF-GIF chimeric protein improves the regeneration efficiency of transgenic plants. Nat Biotechnol **38**: 1274–1279

Deng M, Zhang X, Luo J, Liu H, Wen W, Luo H, Yan J, Xiao Y (2020) Metabolomic analysis reveals differences in evolution between maize and rice. Plant J **103**: 1710–1722

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell **127**: 1309–1321

Domínguez M, Dugas E, Benchouaia M, Leduque B, Jiménez-Gómez JM, Colot V, Quadrana L (2020) The impact of transposable elements on tomato diversity. Nat Commun **11**: 4058

Drewnowski A, Gomez-Carneros C (2000) Bitter taste, phytonutrients, and the consumer: a review. Am J Clin Nutr **72**: 1424–1435

Ekar JM, Betts KJ, Herman AC, Stupar RM, Wyse DL, Brandvain Y, Kantar MB (2019) Domestication in real time: the curious case of a trigenomic sunflower population. Agronomy (Basel) **9**

Emmrich PMF, Rejzek M, Hill L, Brett P, Edwards A, Sarkar A, Field RA, Martin C, Wang TL (2019) Linking a rapid throughput plate-assay with high-sensitivity stable-isotope label LCMS quantification permits the identification and characterisation of low β-L-ODAP grass pea lines. BMC Plant Biol **19**: 489

Endo M, Mikami M, Toki S (2015) Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. Plant Cell Physiol **56**: 41–47

Fedorova M, Herman RA (2020) Obligatory metabolomic profiling of gene-edited crops is risk disproportionate. Plant J **103**: 1985–1988

Fernie AR, Yan J (2019) De novo domestication: an alternative route toward new crops for the future. Mol Plant **12**: 615–631

Ferrero-Serrano Á, Assmann SM (2019) Phenotypic and genome-wide association with the local environment of Arabidopsis. Nat Ecol Evol **3**: 274–285

Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science **289**: 85–88

Fraser PD, Aharoni A, Hall RD, Huang S, Giovannoni JJ, Sonnewald U, Fernie AR (2020) Metabolomics should be deployed in the identification and characterization of gene-edited crops. Plant J **102**: 897–902

Frydman A, Liberman R, Huhman DV, Carmeli-Weissberg M, Sapir-Mir M, Ophir R, Summer LW, Eyal Y (2013) The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. Plant J **73**: 166–178

Gago J, Estrany J, Estes L, Fernie AR, Alorda B, Brotman Y, Flexas J, Escalona JM, Medrano H (2020) Nano and micro unmanned aerial vehicles (UAVs): a new grand challenge for precision agriculture? Curr Protoc Plant Biol **5**: e20103

Gepts P (2014) The contribution of genetic and genomic approaches to plant domestication studies. Curr Opin Plant Biol **18**: 51–59

Grimm DG, Roqueiro D, Salomé PA, Kleeberger S, Greshake B, Zhu W, Liu C, Lippert C, Stegle O, Schölkopf B, et al. (2017) easyGWAS: a cloud-based platform for comparing the results of genome-wide association studies. Plant Cell **29**: 5–19

Gross BL, Olsen KM (2010) Genetic perspectives on crop domestication. Trends Plant Sci **15**: 529–537

Gui S, Yang L, Li J, Luo J, Xu X, Yuan J, Chen L, Li W, Yang X, Wu S, et al. (2020) ZEAMAP, a comprehensive database adapted to the maize multi-omics era. iScience **23**: 101241

Guo J, Li C, Zhang X, Li Y, Zhang D, Shi Y, Song Y, Li Y, Yang D, Wang T (2020) Transcriptome and GWAS analyses reveal candidate gene for seminal root length of maize seedlings under drought stress. Plant Sci **292**: 110380

Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, Wiegert-Rininger K, Wood JC, Douches DS, Farre EM, et al. (2017) Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proc Natl Acad Sci USA **114**: E9999–E10008

Hereford J. (2009) A quantitative survey of local adaptation and fitness trade-offs. Am Nat **173**: 579–588

Howe KL, Contreras-Moreira B, De Silva N, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, et al. (2020) Ensembl genomes 2020-enabling non-vertebrate genomic research. Nucleic Acids Res **48**: D689–D695

Hoyos V, Plaza G, Li X, Caicedo AL (2020) Something old, something new: evolution of Colombian weedy rice (*Oryza* spp.) through de novo de-domestication, exotic gene flow, and hybridization. Evol Appl **13**: 1968–1983

Huang BE, George AW (2011) R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. Bioinformatics **27**: 727–729

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R, et al. (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science **341**: 175–179

Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA (2015) Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnol **15**: 16

Jamnadass R, Mumm RH, Hale I, Hendre P, Muchugi A, Dawson IK, Powell W, Graudal L, Yana-Shapiro H, Simons AJ, et al. (2020) Enhancing African orphan crops with genomics. Nat Genet **52**: 356–360

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJ, Ohyanagi H, Mineta K, Michell CT, Saber N, et al. (2017) The genome of *Chenopodium quinoa*. Nature **542**: 307–312.

Jia Q, Zhang J, Westcott S, Zhang XQ, Bellgard M, Lance R, Li C (2009) GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. Funct Integr Genomics **9**: 255–262

Jin S, Zong Y, Gao Q, Zhu Z, Wang Y, Qin P, Liang C, Wang D, Qiu JL, Zhang F, et al. (2019) Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. Science **364**: 292–295

Khan MA, Olsen KM, Sovero V, Kushad MM, Korban SS (2014) Fruit quality traits have played critical roles in domestication of the apple. Plant Genome **7**

Kim B, Dai X, Zhang W, Zhuang Z, Sanchez DL, Lübberstedt T, Kang Y, Udvardi MK, Beavis WD, Xu S, et al. (2018) GWASpro: a high-performance genome-wide association analysis server. Bioinformatics **35**: 2512–2514

Koenig D, Jimenez-Gomez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, et al. (2013) Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc Natl Acad Sci USA **110**: E2655–E2662

Kumagai M, Nishikawa D, Kawahara Y, Wakimoto H, Itoh R, Tabei N, Tanaka T, Itoh T (2019) TASUKE+: a web-based platform for exploring GWAS results and large-scale resequencing data. DNA Res **26:** 445–452

Kwon CT, Heo J, Lemmon ZH, Capua Y, Hutton SF, Van Eck J, Park SJ, Lippman ZB (2020) Rapid customization of Solanaceae fruit crops for urban agriculture. Nat Biotechnol **38:** 182–188

Lemmon ZH, Reem NT, Dalrymple J, Soyk S, Swartwood KE, Rodríguez-Leal D, Van Eck J, Lippman ZB (2018) Rapid improvement of domestication traits in an orphan crop by genome editing. Nat Plants **4:** 766–770

Lenser T, Theissen G (2013) Molecular mechanisms involved in convergent crop domestication. Trends Plant Sci **18:** 704–714

Li S, Xia L (2020) Precise gene replacement in plants through CRISPR/Cas genome editing technology: current status and future perspectives. aBIOTECH **1:** 58–73

Li T, Yang X, Yu Y, Si X, Zhai X, Zhang H, Dong W, Gao C, Xu C (2018) Domestication of wild tomato is accelerated by genome editing. Nat Biotechnol **36:** 1160+

Lin Z, Li X, Shannon LM, Yeh C-T, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, et al. (2012) Parallel domestication of the Shattering1 genes in cereals. Nat Genet **44:** 720–U154

Liu HJ, Jian L, Xu J, Zhang Q, Zhang M, Jin M, Peng Y, Yan J, Han B, Liu J, et al. (2020) High-throughput CRISPR/Cas9 mutagenesis streamlines trait gene identification in maize. Plant Cell **32:** 1397–1413

Liu J, Fernie AR, Yan J (2021) Crop breeding—from experience-based selection to precision design. J Plant Physiol **256:** 15331

Liu J, Li J, Wang H, Yan J (2020) Application of deep learning in genomics. Sci China Life Sci **63:** 1860–1878

Lotan T, Ohto M, Yee KM, West MA, Lo R, Kwong RW, Yamagishi K, Fischer RL, Goldberg RB, Harada JJ (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell **93:** 1195–1205

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol **24:** 1031–1046

Lu Y, Ye X, Guo R, Huang J, Wang W, Tang J, Tan L, Zhu JK, Chu C, Qian Y (2017) Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. Mol Plant **10:** 1242–1245

Luo C, Fernie AR, Yan J (2020) Single-cell genomics and epigenomics: technologies and applications in plants. Trends Plant Sci **25:** 1030–1040.

Luo J, Wei C, Liu H, Cheng S, Xiao Y, Wang X, Yan J, Liu J (2020) MaizeCUBIC: a comprehensive variation database for a maize synthetic population. Database (Oxford) **2020**

Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, Wang B, Yang Z, Li H, Lin Y, et al. (2015) A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. Mol Plant **8:** 1274–1284

Maher MF, Nasti RA, Vollbrecht M, Starker CG, Clark MD, Voytas DF (2020) Plant gene editing through de novo induction of meristems. Nat Biotechnol **38:** 84–89

Mai CD, Phung NT, To HT, Gonin M, Hoang GT, Nguyen KL, Do VN, Courtois B, Gantet P (2014) Genes controlling root development in rice. Rice (N Y) **7:** 30

McGinn M, Phippen WB, Chopra R, Bansal S, Jarvis BA, Phippen ME, Dorn KM, Esfahanian M, Nazarenus TJ, Cahoon EB, et al. (2019) Molecular tools enabling pennycress (*Thlaspi arvense*) as a model plant and oilseed cash cover crop. Plant Biotechnol J **17:** 776–788

Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J **3:** 269–283

Meng X, Yu H, Zhang Y, Zhuang F, Song X, Gao S, Gao C, Li J (2017) Construction of a genome-wide mutant library in rice using CRISPR/Cas9. Mol Plant **10:** 1238–1241

Meyer RC, Witucka-Wall H, Becher M, Blacha A, Boudichevskaia A, Dörmann P, Fiehn O, Friedel S, von Korff M, Lisec J, et al. (2012) Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. Plant J **71:** 669–683

Meyer RS, DuVal AE, Jensen HR (2012) Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. New Phytol **196:** 29–48

Meyer RS, Puruggganan MD (2013) Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet **14:** 840–852

Nijveen H, Ligterink W, Keurentjes JJB, Loudet O, Long J, Sterken MG, Prins P, Hilhorst HW, de Ridder D, Kammenga JE, et al. (2017) AraQTL—workbench and archive for systems genetics in *Arabidopsis thaliana*. Plant J **89:** 1225–1235

Notaguchi M, Kurotani KI, Sato Y, Tabata R, Kawakatsu Y, Okayasu K, Sawai Y, Okada R, Asahina, M, Ichihashi Y, et al. (2020) Cell–cell adhesion in plant grafting is facilitated by β-1,4-glucanases. Science **369:** 698–702

Obata T, Klemens PAW, Rosado-Souza L, Schlereth A, Gisel A, Stavolone L, Zierer W, Morales N, Mueller LA, Zeeman SC, et al. (2020) Metabolic profiles of six African cultivars of cassava (*Manihot esculenta* Crantz) highlight bottlenecks of root yield. Plant J **102:** 1202–1219

Obata T, Witt S, Lisec J, Palacios-Rojas N, Florez-Sarasa I, Yousfi S, Araus JL, Cairns JE, Fernie AR (2015) Metabolite profiles of maize leaves in drought, heat, and combined stress field trials reveal the relationship between metabolism and grain yield. Plant Physiol **169:** 2665–2683

Pacher M, Puchta H (2017) From classical mutagenesis to nuclease-based breeding—directing natural DNA repair for a natural end-product. Plant J **90:** 819–833

Paterson AH, Lin YR, Li ZK, Schertz KF, Doebley JF, Pinson SRM, Liu SC, Stansel JW, Irvine JE (1995a) Convergent domestication of cereal crops by independent mutations at corresponding genetic-loci. Science **269:** 1714–1718

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL (1995b) The weediness of wild plants—molecular analysis of genes influencing dispersal and persistence of Johnsongrass, *Sorghum-halepense* (l) pers. Proc Natl Acad Sci USA **92:** 6127–6131

Pavlidis P, Alachiotis N (2017) A survey of methods and tools to detect recent and strong positive selection. J Biol Res (Thessalon) **24:** 7

Pearce S, Vazquez-Gross H, Herin SY, Hane D, Wang Y, Gu YQ, Dubcovsky J (2015) WheatExp: an RNA-seq expression database for polyploid wheat. BMC Plant Biol **15:** 299

Price EJ, Drapal M, Perez-Fons L, Amah D, Bhattacharjee R, Heider B, Rouard M, Swennen R, Becerra Lopez-Lavalle LA, et al. (2020) Metabolite database for root, tuber, and banana crops to facilitate modern breeding in understudied crops. Plant J **101:** 1258–1268

Qi W, Zhu T, Tian Z, Li C, Zhang W, Song R (2016) High-efficiency CRISPR/Cas9 multiplex gene editing using the glycine tRNA-processing system-based strategy in maize. BMC Biotechnol **16:** 58

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nat Genet **44:** 217–220

Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB (2017) Engineering quantitative trait variation for crop improvement by genome editing. Cell **171:** 470–480

Rosado-Souza L, Scossa F, Chaves IS, Kleessen S, Salvador LF, Milagre JC, Finger F, Bhering LL, Sulpice R, Araújo WL, et al. (2015) Exploring natural variation of photosynthetic, primary metabolism and growth parameters in a large panel of *Capsicum chinense* accessions. Plants **242:** 677–691

Scarcelli N, Cubry P, Akakpo R, Thuillet AC, Obidiegwu J, Baco MN, Otoo E, Sonke B, Dansi A, Djedatin G, et al. (2019) Yam

genomics supports West Africa as a major cradle of crop domestication. Sci Adv **5**: eaaw1947

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, et al. (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol **24**: 447–454

Schmidt C, Fransz P, Rönspies M, Dreissig S, Fuchs J, Heckmann S, Houben A, Puchta H (2020) Changing local recombination patterns in Arabidopsis by CRISPR/Cas mediated chromosome engineering. Nat Commun **11**: 4418

Schneeberger K, Weigel D (2011) Fast-forward genetics enabled by new sequencing technologies. Trends Plant Sci **16**: 282–288

Schwahn K, de Souza LP, Fernie AR, Tohge T (2014) Metabolomics-assisted refinement of the pathways of steroidal glycoalkaloid biosynthesis in the tomato clade. J Integr Plant Biol **56**: 864–875

Schwartz C, Lenderts B, Feigenbutz L, Barone P, Llaca V, Fengler K, Svitashev S (2020) CRISPR-Cas9-mediated 75.5-Mb inversion in maize. Nat Plants **6**: 1427–1431

Seren Ü, Vilhjálmsson BJ, Horton MW, Meng D, Forai P, Huang YS, Long Q, Segura V, Nordborg M (2012) GWAPP: a web application for genome-wide association mapping in Arabidopsis. Plant Cell **24**: 4793–4805

Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W, et al. (2014) Plant science. Biosynthesis, regulation, and domestication of bitterness in cucumber. Science **346**: 1084–1088

Sheehan MJ, Kennedy LM, Costich DE, Brutnell TP (2007) Subfunctionalization of PhyB1 and PhyB2 in the control of seedling and mature plant traits in maize. Plant J **49**: 338–353

Smith JM, Haigh J (2007) The hitch-hiking effect of a favourable gene. Genet Res **89**: 391–403

Smýkal P, Trněný O, Brus J, Hanáček P, Rathore A, Roma RD, Pechanec V, Duchoslav M, Bhattacharyya D, Bariotakis M, et al. (2018) Genetic structure of wild pea (*Pisum sativum* subsp. elatius) populations in the northern part of the Fertile Crescent reflects moderate cross-pollination and strong effect of geographic but not environmental distance. PLoS ONE **13**: e0194056

Sonnewald U, Fernie AR, Gruissem W, Schläpfer P, Anjanappa RB, Chang SH, Ludewig F, Rascher U, Muller O, van Doorn AM, Rabbi IY, et al. (2020) The Cassava Source-Sink project: Opportunities and challenges for crop improvement by metabolic engineering. Plant J **103**: 1655–1665

Sork VS (2018) Genomic studies of local adaptation in natural plant populations. J Hered **109**: 3–15

Sosso D, Luo D, Li Q-B, Sasse J, Yang J, Gendrot G, Suzuki M, Koch KE, McCarty DR, Chourey PS, et al. (2015) Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. Nat Genet **47**: 1489+

Spannagl M, Nussbaumer T, Bader KC, Martis MM, Seidel M, Kugler KG, Gundlach H, Mayer KF (2016) PGSB PlantsDB: updates to the database framework for comparative plant genome research. Nucleic Acids Res **44**: D1141–1147

Stetter MG, Gates DJ, Mei W, Ross-Ibarra J (2018). How to make a domesticate. Curr Biol **27**: R896–R900

Stetter MG, Vidal-Villarejo M, Schmid KJ (2020) Parallel seed color adaptation during multiple domestication attempts of an ancient new world grain. Mol Biol Evol **37**: 1407–1419

Stetter MG (2020) Limits and constraints to crop domestication. Am J Bot **107**: 1617–1621

Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proc Natl Acad Sci USA **98**: 11806–11811

Studer AJ, Wang H, Doebley JF (2017) Selection during maize domestication targeted a gene network controlling plant and inflorescence architecture. Genetics **207**: 755–765

Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM (2012) Reshaping of the maize transcriptome by domestication. Proc Natl Acad Sci USA **109**: 11878–11883

Takahashi Y, Sakai H, Yoshitsu Y, Muto C, Anai T, Pandiyan M, Senthil N, Tomooka N, Naito K (2019) Domesticating *Vigna Stipulacea*: A potential legume crop with broad resistance to biotic stresses. Front Plant Sci **10**: 1607

Tello-Ruiz MK, Naithani S, Stein JC, Gupta P, Campbell M, Olson A, Wei S, Preece J, Geniza MJ, Jiao Y, et al. (2018) Gramene 2018: unifying comparative genomics and pathway resources for plant research. Nucleic Acids Res **46**: D1181–D1189

Tian H, Xu R, Canadell JG, Thompson RL, Winiwarter W, Suntharalingam P, Davidson EA, Ciais P, Jackson RB, Janssens-Maenhout G, et al. (2020) A comprehensive quantification of global nitrous oxide sources and sinks. Nature **586**: 248–256

Togninalli M, Seren Ü, Meng D, Fitz J, Nordborg M, Weigel D, Borgwardt K, Korte A, Grimm DG (2018) The AraGWAS catalog: a curated and standardized *Arabidopsis thaliana* GWAS catalog. Nucleic Acids Res **46**: D1150–D1156

Tohge T, Scossa F, Fernie AR (2015) Integrative approaches to enhance understanding of plant metabolic pathway structure and regulation. Plant Physiol **169**: 1499–1511

Tohge T, Scossa F, Wendenburg R, Frasse P, Balbo I, Watanabe M, Alseekh S, Jadhav SS, Delfin JC, Lohse M, et al. (2020) Exploiting natural variation in tomato to define pathway structure and metabolic regulation of fruit polyphenolics in the *lycopersicum* complex. Mol Plant **13**: 1027–1036

Tong H, Küken A, Nikoloski Z (2020) Integrating molecular markers into metabolic models improves genomic selection for Arabidopsis growth. Nat Commun **11**: 2410

Tong H, Nikoloski Z (2020) Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data. J Plant Physiol **257**: 153354

Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science **310**: 1031–1034

Uga Y, Sugimoto K, Ogawa S, Rane J, Ishitani M, Hara N, Kitomi Y, Inukai Y, Ono K, Kanno N, et al. (2013) Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. Nat Genet **45**: 1097–1102

Van Tassel DL, Tesdell O, Schlautman B, Rubin MJ, DeHaan LR, Crews TE, Streit Krug A (2020) New food crop domestication in the age of gene editing: genetic, agronomic and cultural change remain co-evolutionarily entangled. Front Plant Sci **11**: 789

Vergara-Diaz O, Vatter T, Kefauver SC, Obata T, Fernie AR, Araus JL (2020) Assessing durum wheat ear and leaf metabolomes in the field through hyperspectral data. Plant J **102**: 615–630

Vergara-Diaz O, Vatter T, Vicente R, Obata T, Nieto-Taladriz MT, Aparicio N, Carlisle Kefauver S, Fernie A, Araus JL (2020) Metabolome profiling supports the key role of the spike in wheat yield performance. Cells **9**

Wang J, Qi M, Liu J, Zhang Y (2015) CARMO: a comprehensive annotation platform for functional exploration of rice multi-omics data. Plant J **83**: 359–374

Wang M, Li W, Fang C, Xu F, Liu Y, Wang Z, Yang R, Zhang M, Liu S, Lu S, et al. (2018) Parallel selection on a dormancy gene during domestication of crops from multiple families. Nat Genet **50**: 1435–

Wang W, Feng B, Xiao J, Xia Z, Zhou X, Li P, Zhang W, Wang Y, Møller BL, Zhang P, et al. (2014) Cassava genome from a wild ancestor to cultivated varieties. Nat Commun **5**: 5110

Watt M, Fiorani F, Usadel B, Rascher U, Muller O, Schurr U (2020) Phenotyping: new windows into the plant for breeders. Annu Rev Plant Biol **71:** 689–712

Wei X, Qui J, Yong K, Fan J, Zhang Q, Hua H, Liu J, Wang Q, Olsen KM, Han B, **et al.** (2021) A quantitative genomics map of rice provides genetic insights and guides breeding. Nat Genet **53:** 243–253

Wilkinson PA, Winfield MO, Barker GLA, Allen AM, Burridge A, Coghill JA, Edwards KJ (2012) CerealsDB 2.0: an integrated resource for plant breeders and scientists. BMC Bioinformatics **13:** 219

Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. Science **319:** 1527–1530

Xie K, Yang Y (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. Mol Plant **6:** 1975–1983

Xu G, Cao J, Wang X, Chen Q, Jin W, Li Z, Tian F (2019) Evolutionary metabolomics identifies substantial metabolic divergence between maize and its wild ancestor, Teosinte. Plant Cell **31:** 1990–2009

Yamagishi N, Kishigami R, Yoshikawa N (2014) Reduced generation time of apple seedlings to within a year by means of a plant virus vector: a new plant-breeding technique with no transmission of genetic modification to the next generation. Plant Biotechnol J **12:** 60–68

Yang J, Moeinzadeh MH, Kuhl H, Helmuth J, Xiao P, Haas S, Liu G, Zheng J, Sun Z, Fan W, **et al.** (2017) Haplotype-resolved sweet potato genome traces back its hexaploidization history. Nat Plants **3:** 696–703

Yang W, Feng H, Zhang X, Zhang J, Doonan JH, Batchelor WD, Xiong L, Yan J (2020) Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. Mol Plant **13:** 187–214

Ye J, Tian R, Meng X, Tao P, Li C, Liu G, Chen W, Wang Y, Li H, Ye Z, **et al.** (2020) Tomato SD1, encoding a kinase interacting protein, is a major locus controlling stem development. J Exp Bot **71:** 3575–3587

Youens-Clark K, Buckler E, Casstevens T, Chen C, Declerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, **et al.** (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res **39:** D1085–D1094

Yu H, Lin T, Meng X, Du H, Zhang J, Liu G, Chen M, Jing Y, Kou L, Li X, **et al.** (2021) A route to de novo domestication of wild allotetraploid rice. Cell **184**: 1156–1170.e14

Zhang H, Li Y, Zhu J-K (2018) Developing naturally stress-resistant crops for a sustainable agriculture. Nat Plants **4:** 989–996

Zhang H, Zhang J, Wei P, Zhang B, Gou F, Feng Z, Mao Y, Yang L, Zhang H, Xu N, **et al.** (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. Plant Biotechnol J **12:** 797–807

Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, **et al.** (2018) Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nat Genet **50:** 1565–

Zhang L, Su W, Tao R, Zhang W, Chen J, Wu P, Yan C, Jia Y, Larkin RM, Lavelle D, **et al.** (2017) RNA sequencing provides insights into the evolution of lettuce and the regulation of flavonoid biosynthesis. Nat Commun **8:** 2264

Zhang W, Alseekh S, Zhu X, Zhang Q, Fernie AR, Kuang H, Wen W (2020) Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. Plant J **104:** 613–630

Zhang W, Zhang Y, Qiu H, Guo Y, Wan H, Zhang X, Scossa F, Alseekh S, Zhang Q, Wang P, **et al.** (2020) Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties. Nat Commun **11:** 3719

Zhao H, Yao W, Ouyang Y, Yang W, Wang G, Lian X, Xing Y, Chen L, Xie W (2014) RiceVarMap: a comprehensive database of rice genomic variations. Nucleic Acids Res **43:** D1018–D1022

Zhou Y, Ma Y, Zeng J, Duan L, Xue X, Wang H, Lin T, Liu Z, Zeng K, Zhong Y, **et al.** (2016) Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. Nat Plants **2:** 16183

Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, Lin T, Qin M, Peng M, Yang C, **et al.** (2018) Rewiring of the fruit metabolome in tomato breeding. Cell **172:** 249–261.e212

Zsögön A, Čermák T, Naves ER, Notini MM, Edel KH, Weinl S, Freschi L, Voytas DF, Kudla J, Peres LEP (2018) De novo domestication of wild tomato using genome editing. Nat Biotechnol **36:** 1211–1216

Zuo J, Niu QW, Frugis G, Chua NH (2002) The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. Plant J **30:** 349–359